

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-278706

(43)Date of publication of application : 27.09.2002

(51)Int.Cl.

G06F 3/06

G06F 12/16

(21)Application number : 2001-081875

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 22.03.2001

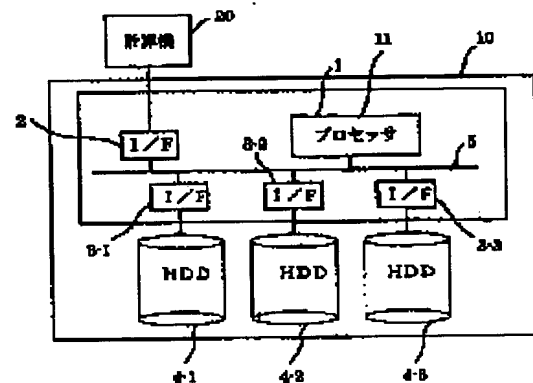
(72)Inventor : ARAI SEIJI
ICHISE ATSUSHI

(54) DISK ARRAY DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a disk array device by which a backup processing of data is enabled without lowering performance.

SOLUTION: When a disk controller 11 receives a backup request, it integrates the same piece of data is stored and one HDD 4-2 having duplexed structure in a backup disk and integrates a HDD 4-3 of a backup disk in the duplexed structure with a HDD 4-1 as a master disk. And copy of the data to the HDD 4-3 is performed from the HDD 4-2 and when a new writing request is received, the data is written in the HDDs 4-1 and 4-3.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

W1147

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-278706

(P2002-278706A)

(43) 公開日 平成14年9月27日 (2002.9.27)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード(参考)
G 0 6 F 3/06	3 0 4	G 0 6 F 3/06	3 0 4 F 5 B 0 1 8
	5 4 0		5 4 0 5 B 0 6 5
12/16	3 1 0	12/16	3 1 0 M
	3 2 0		3 2 0 L

審査請求 未請求 請求項の数 3 O L (全 7 頁)

(21) 出願番号 特願2001-81875 (P2001-81875)

(22) 出願日 平成13年3月22日 (2001.3.22)

(71) 出願人 000003078

株式会社東芝

東京都港区芝浦一丁目1番1号

(72) 発明者 新井 誠二

東京都府中市東芝町1番地 株式会社東芝
府中事業所内

(72) 発明者 市瀬 敦司

東京都府中市東芝町1番地 株式会社東芝
府中事業所内

(74) 代理人 100083161

弁理士 外川 英明

Fターム(参考) 5B018 GA04 HA04 MA14 QA20

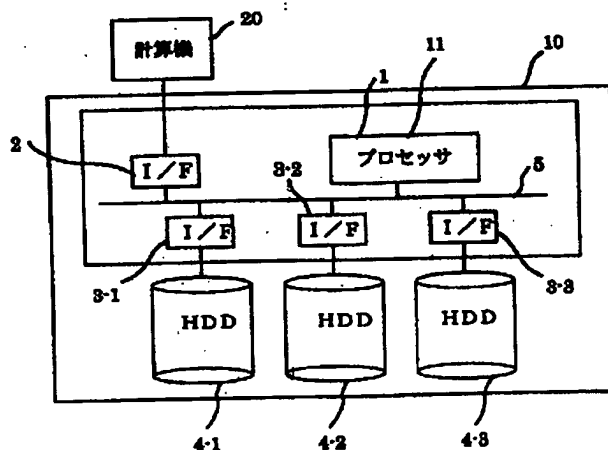
5B065 BA01 CA11 CA30 EA12 EA35

(54) 【発明の名称】 ディスクアレイ装置

(57) 【要約】

【課題】 性能を低下させることなくデータのバックアップ処理が可能なディスクアレイ装置を提供する。

【解決手段】 ディスクコントローラ11がバックアップ要求を受信すると、同一のデータを記憶し2重化構成をとっている一方のHDD4-2をバックアップディスクに、そして予備ディスクのHDD4-3をマスタディスクとしてHDD4-1との2重化構成に組み込む。そして、このHDD4-3に対するデータのコピーはHDD4-2から行うとともに、新たに書き込み要求を受けた場合には、HDD4-1とHDD4-3に書き込む。



【特許請求の範囲】

【請求項1】 複数のハードディスクと、各ハードディスクに対するデータの書き込み及び読出しを制御するディスクコントローラとを備えたディスクアレイ装置において、

前記複数のハードディスクが、

それぞれが少なくとも1台のハードディスクを有し、同一のデータを記憶する第1及び第2のディスクグループと、少なくとも1台のハードディスクを有し予備用に設定された第3のディスクグループとを構成し、

前記ディスクコントローラは、データのバックアップ要求を受信すると、

前記第2のディスクグループに記憶されているデータを前記第3のディスクグループにコピーし、

データの書き込みは、前記第1のディスクグループ及び第3のディスクグループへ行うことを特徴とするディスクアレイ装置。

【請求項2】 前記第1のディスクグループから第3のディスクグループへのコピー処理は、データを保持するブロックと該ブロックに対するコピー完了及び未完了を示す情報からなるテーブルによって管理されることを特徴とする請求項1に記載のディスクアレイ装置。

【請求項3】 データのバックアップ処理中に前記ディスクコントローラがデータの書き込み要求を受信すると、前記テーブルを参照してコピーが完了しているか確認し、

コピーが未完了の場合には、前記第1のディスクグループと第3のディスクグループにデータの書き込みを行うとともに、前記テーブル内の前記情報をコピー完了に設定することを特徴とする請求項2に記載のディスクアレイ装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数のディスクドライブを備えたディスクアレイ装置に関し、特に、前記複数のディスクドライブを2つのグループに分け、同一のデータを2重化して記憶することが可能なディスクアレイ装置に関する。

【0002】

【従来の技術】複数のディスクドライブ（以下、HDDと称する。）を備えたディスクアレイ装置では、これら複数のHDDを2つのグループに分け、双方のグループに同一のデータを記憶する構成を取るものがあり、この技術は、データに冗長性を持たせる方式を定義したRAID (Redundant Arrays of Inexpensive Disks) と呼ばれる技術の中で、RAID1として定義されている。そして、データを2重化して記憶するディスクアレイ装置は、一方のグループを構成するHDDに読出し不能な故障が発生した場合でも、他方のグループに同一のデータが記憶されていることから、データの消失を回避

することが可能であり、また故障したHDDを交換して正常なグループのHDDからデータをコピーすることによって、再び2重化構成を形成することが可能である。しかしながら、このように信頼性の高いディスクアレイ装置であっても、定期的にデータのバックアップを外部の記憶媒体に取る場合があり、上述のRAID1の方式を採用しているディスクアレイ装置では、ディスクアレイ装置に組み込まれているHDDの他にバックアップデータを記憶するための磁気テープ等の記憶媒体を用意し、一方のグループのHDDからデータを読み出してこの記憶媒体にコピーを取るによりバックアップデータの作成を行っている。

【0003】そしてこのとき、ディスクアレイ装置の信頼性、つまりデータが2重化された状態を維持するために、ネットワークを介して接続された計算機から書き込み要求を受けると、この書き込み要求にかかるデータが書き込まれる領域が既にバックアップ用のHDDや磁気テープ等の記憶媒体にコピーされているか確認し、コピーが完了している場合には双方のグループにデータを書込み、他方、コピーが未完了の場合にはミラーリングを構成する一方のグループからデータを読み出し、バックアップ用の記憶媒体にコピーしてから双方のグループに対してデータの書き込みを行っている。

【0004】

【発明が解決しようとする課題】しかしながら、上述のようなRAID1を採用したディスクアレイ装置におけるデータのバックアップ処理にも、まだ解決すべき課題が残されている。即ち、上述の方式では、データのバックアップ処理中に計算機から書き込み要求を受けると、書き込みデータが書き込まれる領域に記憶されているデータのコピー処理が完了していなければ、この領域に記憶されているデータを一旦読出してバックアップ用の記憶媒体へコピーしてから書き込み処理を行うため、バックアップ処理中にコピーの未完了の領域に対する書き込み要求が多数発生した場合には、ディスクアレイ装置の性能が低下してしまうといった欠点がある。そこで本発明では、上記欠点を解消し、信頼性を維持するとともにディスクアレイ装置の性能を低下させることなくバックアップ処理が可能なディスクアレイ装置を提供することを目的とする。

【0005】

【課題を解決するための手段】上記課題を解決するため、請求項1にかかるディスクアレイ装置では、複数のハードディスクと、各ハードディスクに対するデータの書き込み及び読出しを制御するディスクコントローラとを備えたディスクアレイ装置において、前記複数のハードディスクが、それぞれが少なくとも1台のハードディスクを有し同一のデータを記憶する第1及び第2のディスクグループと、少なくとも1台のハードディスクを有し予備用に設定された第3のディスクグループとを構成

し、前記ディスクコントローラは、データのバックアップ要求を受信すると、前記第2のディスクグループに記憶されているデータを前記第3のディスクグループにコピーし、データの書き込みは、前記第1のディスクグループ及び第3のディスクグループへ行うことを特徴としている。このように構成することにより、第2のディスクグループから第3のディスクグループに対するコピー処理中であっても、この処理の状況に影響されることがなくデータの書き込み要求にかかる処理を実行することが可能となる。また、請求項2にかかるディスクアレイ装置では、請求項1のディスクアレイ装置において、前記第1のディスクグループから第3のディスクグループへのコピー処理は、データを保持するブロックと該ブロックに対するコピー完了及び未完了を示す情報からなるテーブルによって管理されることを特徴としている。

【0006】このように構成することにより、第2のディスクグループから第3のディスクグループに対するコピー処理中であっても、同様に、この処理の状況に影響されることがなくデータの書き込み要求にかかる処理を実行することが可能となる。また、請求項3にかかるディスクアレイ装置では、請求項2のディスクアレイ装置において、データのバックアップ処理中に前記ディスクコントローラがデータの書き込み要求を受信すると、前記テーブルを参照してコピーが完了しているか確認し、コピーが未完了の場合には、前記第1のディスクグループと第3のディスクグループにデータの書き込みを行うとともに、前記テーブル内の前記情報をコピー完了に設定することを特徴としている。このように構成することにより、第2のディスクグループから第3のディスクグループに対するコピー処理中であっても、この処理の状況に影響されることがなくデータの書き込み要求にかかる処理を実行することが可能となり、さらにデータのコピー処理中にコピーが未完了の領域に対してデータの書き込み要求を実行した場合に、コピーが完了している旨が設定可能であるため、上記作用に加えコピー処理によって古いデータが上書きされることを防止できる。

【0007】

【発明の実施の形態】以下、添付の図面を参照して本発明の実施の形態について説明する。図1は、ディスクアレイ装置10の概略を示す図であり、このディスクアレイ装置10はディスクコントローラ11と3台のHDD 4-1～4-3を有し、さらにディスクコントローラ11は、プロセッサ1、外部インタフェース2、ディスクインタフェース3-1～3-3、そして内部バス5とを備えている。ディスクコントローラ11はディスクアレイ装置10を制御するもので、外部機器からのデータ書き込み要求を受けると、このデータを固定長のデータ（ストライピングデータ）に分割して、複数のHDDに対して分散して記憶するとともに、読み出し要求をうけると、複数のHDDからストライピングデータを読み出

して元のデータに復元し、外部機器へ送信する機能、いずれかのHDDに故障が発生した場合に、このHDDに記憶されていたデータを復元する機能等を担っている。HDD 4-1～HDD 4-3は、ディスクコントローラ11によって制御され、外部機器で使用されるデータをストライピングデータとして記憶するもので、各HDDは、図2に示すようにブロックサイズ等のデータ管理に必要な情報を記憶するための管理情報領域41と、データを記憶するためのデータ領域42とから構成されており、更にデータ領域42は固定長に分割されたデータを記憶するために、複数のブロック42a、42b、42c、・・・から構成されている。

【0008】なお、本実施の形態では、ディスクアレイ装置10がRAID1に準拠しているため、2台のHDDが同一のデータを記憶し、他の1台が予備ディスクとして扱われる。プロセッサ1は、ディスクコントローラ11の制御を行うもので、上述のディスクコントローラ11の機能はこのプロセッサ1によって実現される。外部インタフェース2は、外部機器からデータの書き込み及び読み出し要求を受信するとともに、この外部機器に対して各種データを送信する。ディスクインタフェース3-1～3-3は、それぞれHDD 4-1～HDD 4-3に対応して設けられており、プロセッサ1の要求に基づいて、ストライピングデータの書き込み及び読み出しを行う。そして、内部バス5は、プロセッサ1、外部インタフェース2、及びディスクインタフェース3-1～3-3を接続するとともに、各モジュールによって各種データの送受信に利用される。なお、図1では説明を簡単にするために3台のHDDを備えたディスクアレイ装置を例にとっているが、これ以上の数からなるHDDを備えていても構わない。ただし、この場合には同じ台数のHDDから成るグループを3つ用意する必要があるため、ディスクアレイ装置は3の倍数のHDDを備えることが好ましい。

【0009】また、ディスクアレイ装置10には、外部記憶装置として計算機20がネットワークを介して接続されており、読み出し／書き込み要求の他、バックアップ処理の開始／終了要求をディスクアレイ装置10に対して発行する。つづいて、上述の構成を有するディスクアレイ装置10の動作について説明するが、ディスクアレイ装置10が起動された際の初期状態では、HDD 4-1とHDD 4-2が2重化の構成をとって同一のデータを記憶し、HDD 4-3が予備ディスクであるものとする。まず、電源がオンにされて、ディスクチェック等の起動処理が完了すると、ディスクアレイ装置10は計算機20によって使用可能となり、計算機20によって発行されるデータの書き込み要求や読み出し要求の受信、および計算機20に対してデータの送信が行える状態になる。このような起動処理の完了後、計算機20によってデータの書き込み要求が発行されると、この要求は外

部インタフェース2を介してディスクコントローラ11に取り込まれる。取り込まれた書き込み要求は、プロセッサ1が管理するキャッシュメモリ（図示せず）に書き込まれた後にプロセッサ1によって受信した要求の種別が判定される。

【0010】ここでは上述の通り書き込み要求であることから、この要求に含まれたデータは複数のストライピングデータに分割され、同一のストライピングデータがディスクインタフェース3-1と3-2を介してHDD4-1及びHDD4-2内の同一位置にあるデータ領域に記憶される。また、計算機20から発行された要求がデータの読み出し要求であった場合、プロセッサ1は、ディスクインタフェース3-1または3-2のいずれかを介してHDD4-1または4-2よりストライピングデータを読み出し、元のデータに復元した後に外部インタフェースを介して計算機20へ送信する。なお、ストライピングデータを読み出す際、ディスクヘッドの位置からより短時間でストライピングデータを読み出せるHDDを選択している。このように、計算機20によって発行されるデータの書き込み及び読み出し処理を実行しているときに、計算機20からデータのバックアップ要求を受信すると、ディスクアレイ装置10は図3および図4で示される流れに従ってデータのバックアップ処理を実行する。プロセッサ1は、計算機20によって発行され外部インタフェース2を介して受信する要求を常に監視しており（ステップ1）、データの書き込み・読み出し要求といった通常の要求である場合には、この要求で指定された処理を実行する（ステップ2）。

【0011】一方、プロセッサ1が受信した要求がバックアップ要求であった場合、2重化を構成する一方のHDD4-2をバックアップディスクに、そして予備ディスクであるHDD4-3をHDD4-1と2重化を構成するよう設定変更を行う（ステップ3）。以下、バックアップ処理中において、2重化を構成する2台のHDDをマスタディスク、そしてバックアップデータを保持するHDDをバックアップディスクと呼ぶ。即ち、上述のようにバックアップ処理が開始されるとHDD4-1とHDD4-3がマスタディスクとなり、HDD4-2がバックアップディスクとなる。次に、HDD4-2をバックアップディスクに、そしてHDD4-3をマスタディスクへと設定が完了すると、プロセッサ1の制御によりバックアップディスクHDD4-2に記憶されているデータは、アドレス番号の小さい方から順にブロック単位でマスタディスクHDD4-3へとコピーされる（ステップ4）。このバックアップディスクHDD4-2から新たなマスタディスクHDD4-3へのコピー処理中、プロセッサ1は計算機20から新たな要求を受信したかを監視し（ステップ5）、新たな要求を受信すると、この要求が書き込み要求であるか読み出し要求であるかの確認を行う（ステップ6）。

【0012】ここで、新たな要求が読み出し要求の場合には、インタフェース3-1を介し、マスタディスクHDD4-1から該当するストライピングデータを読み出し、元のデータに復元処理を行った後に外部インタフェース2を介して計算機20へ送信する（ステップ7）。一方、新たな要求が書き込み要求の場合には、マスタディスクHDD4-3の書き込み先のブロックが既にコピー済みであるか確認し（ステップ8）、コピーが済んでいる場合には、書き込み要求にかかるデータをストライピングデータに分割した後にマスタディスクHDD4-1およびマスタディスクHDD4-3へと書き込む（ステップ9）。そして、コピーが未完了な書き込み領域に対する書き込み要求であった場合には、マスタディスクHDD4-1とマスタディスクHDD4-3へ書き込み要求にかかるデータをストライピングデータに分割して書き込むと共に、書き込み処理を行った書き込み領域の状態をコピー済みに設定する（ステップ10）。ここで、HDDの各ブロックに対しコピーが完了しているか否かの情報は、HDD4-1～4-3のうちの少なくとも1台のHDDの管理情報領域41に、例えば、図5のようなバックアップ状態テーブル50を作成することによって管理される。

【0013】そしてバックアップ状態テーブル50は、ブロック番号とバックアップ状態フラグから構成されており、該フラグが1のときにバックアップ完了を、そしてバックアップ状態フラグが0のときにバックアップ未完了を示している。例えば、バックアップ状態テーブル50にコピー未完了、即ちバックアップ状態フラグが0として記録されているブロック42cに対して書き込みを行う場合、マスタディスクHDD4-1とマスタディスクHDD4-3のブロック42cにデータの書き込みを行った後、バックアップ状態テーブル50のブロック42cに対応するバックアップ状態フラグを0から1に更新する。このとき、バックアップ状態テーブル50は図6に示した状態になっている。以上の処理は、マスタディスク4-2に記憶されているデータと同一のデータが、マスタディスク4-3に形成されるまで実行され（ステップ11）、コピー処理が完了すると、プロセッサ1は計算機20に対してバックアップ処理が完了した旨を通知する。このバックアップ処理が完了した後、各HDDの設定は変更されることなく、HDD4-1とHDD4-3が2重化を構成し、HDD4-2が予備ディスクとして運用される。

【0014】即ち、次のバックアップ要求を受信した場合には、HDD4-3がバックアップディスクに、そしてHDD4-2がHDD4-1と2重化構成をとってマスタディスクとなる。以上説明したように、データのバックアップ処理を実行中に計算機20から書き込み要求を受信した際、この書き込み要求が記憶されているデータ領域のデータがバックアップディスクにコピーされて

いるか否かに影響されることなく書き込み処理を実行できるため、ディスクアレイ装置の性能を低下させることなくバックアップ処理を行うことが可能となる。また、バックアップディスクに記憶されているデータの読み出しが必要な場合でも、バックアップ処理の開始時点でバックアップデータが確保されていることから、バックアップ処理の進捗状況に左右されることなくデータの読み出しを行うことが可能となる。つまり、バックアップディスクにデータがコピーされているか否か、そしてマスタディスクのデータに更新があるか否かによって読み出すHDDを選択する必要がなくなる。なお、上述の説明では新たに設定されたマスタディスクへのコピーは、バックアップディスクからとしたが、全データを記憶しているマスタディスクから行ってもよいし、このマスタディスクとバックアップディスクの双方から選択的行ってもよい。

【0015】また、上述の説明では、バックアップ処理について説明したが、ある時点におけるデータの保存状態を確保するスナップショット機能を実行する場合にも、本発明は適用可能である。

【0016】

【発明の効果】以上説明したように、本発明のディスクアレイ装置によれば、データのバックアップ処理中に計算機から書き込み要求が発行された場合であっても、信

頼性を維持するとともに、ディスクアレイ装置の性能、即ち、書き込み処理に要する時間を遅延することを防止することが可能となる。

【図面の簡単な説明】

【図1】 本発明にかかるディスクアレイ装置の構成を示す図。

【図2】 ハードディスクの構成を示す図。

【図3】 本発明にかかるバックアップ処理を示す図。

【図4】 本発明にかかるバックアップ処理を示す図。

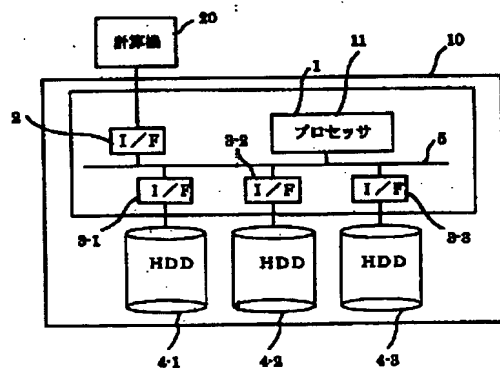
【図5】 バックアップ状態テーブルの構成を示す図。

【図6】 バックアップ状態テーブルの状態を示す図。

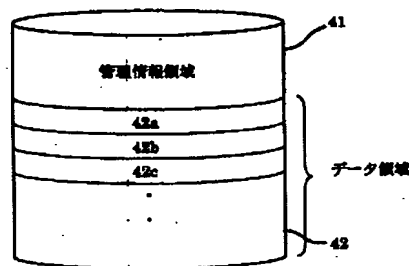
【符号の説明】

- 1 … プロセッサ
- 2 … 外部インタフェース
- 3-1～3-3 … ディスクインタフェース
- 4-1～4-3 … ハードディスク
- 5 … 内部バス
- 10 … ディスクアレイ装置
- 11 … ディスクコントローラ
- 20 … 計算機
- 41 … 管理情報領域
- 42 … データ領域
- 50 … バックアップ状態テーブル

【図1】



【図2】



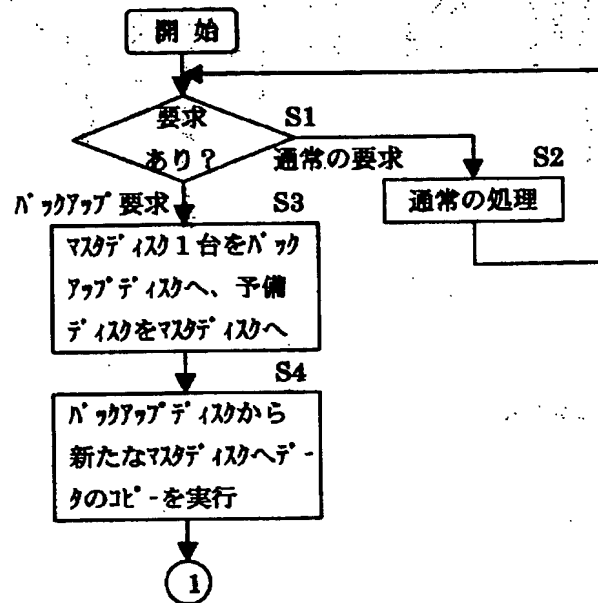
【図6】

データ領域	コピー状態
42a	0
42b	0
42c	1
⋮	⋮

【図5】

データ領域	コピー状態
42a	0
42b	0
42c	0
⋮	⋮

【図3】



【図4】

